

# **ANÁLISIS DE LA EVOLUCIÓN DE LAS BÚSQUEDAS SIMPLES HASTA LA INNOVACIÓN DE LAS BÚSQUEDAS RELACIONADAS, UNA WEB SEMÁNTICA**

***Ariel Valdez Rodríguez***

Tecnológico Nacional de México en Celaya

*13030367@itcelaya.edu.mx*

***Elizabeth Rosas Trilla***

Tecnológico Nacional de México en Celaya

*13030386@itcelaya.edu.mx*

***Jessica Alejandra Araujo Rodríguez***

Tecnológico Nacional de México en Celaya

*jessica\_rodriguez95@outlook.com*

***Rodrigo Daniel Heredia Heredia***

Tecnológico Nacional de México en Celaya

*reaper.daniel@hotmail.com*

## **Resumen**

La web semántica es considerada una rama de la inteligencia artificial y las tecnologías Web, que permite ordenar la información en el motor de búsqueda para evitar así la redundancia, el exceso de información, mejorar las búsquedas, reduce el costo y el tiempo que se invierte al consultar información útil en Internet, además, mantiene una búsqueda coherente que permite que las computadoras puedan procesar la información de manera inteligente. Así mismo, se emplea la búsqueda en lenguaje natural lo que permite que la consulta se enfoque más al significado de lo que se desea conocer y no a los elementos que componen la estructura de la consulta en la web, también utiliza la minería de datos que le permite conocer la información que se encuentra oculta, es decir, toda la información relevante, útil y no evidente para el motor de búsqueda.

**Palabra(s) Clave:** Búsqueda de datos, Inteligencia artificial, Marco de Descripción de Recursos (RDF), Navegadores web.

## **Abstract**

*The semantic web is considered a branch of the artificial intelligence (A.I.) and the web technologies, it allows information sort at engine search to prevent redundancy, too much information, search optimization, cost and time reduction at the look up for relevant information, also, keep it in a coherent search that allows computers to process information on a smartly way. Likewise, it's used to search in a natural language that allow to the look up on a perspective on the meaning of what do you want to know less than the components of the web search structure, also use data mining that allows to know hiding information, it means, all the relevant, useful and not too obvious information, to the search engine.*

**Keywords:** Artificial Intelligence, Data search, Resource Description Framework (RDF), Web.browsers.

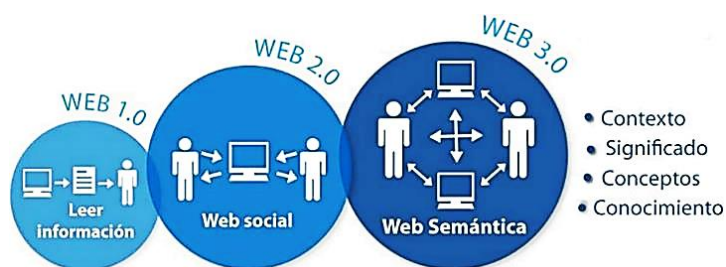
## **1. Introducción**

Al paso del tiempo notaremos como lentamente nos encontramos más inmersos en un mundo que gira alrededor de la web hasta tal punto que se ha convertido en una herramienta de uso habitual en nuestra sociedad en diferentes ámbitos. El desarrollo de la tecnología al inicio del siglo XXI ha demostrado la necesidad de adaptarse velozmente al cambio en las actividades más comunes del ser humano, el contacto con otras personas ha pasado de ser verbal a ser por medio de un chat en una plataforma en la web, las investigaciones escolares se realizan por medio de un motor de búsqueda y aplicaciones en diferentes dispositivos, de igual manera las noticias sobre el mundo ahora se consultan por Internet.

Esto ha llevado a la conclusión de evolucionar el mundo de las consultas en la web que permitan el manejo de la información de manera eficiente, reduciendo el tiempo de consultas y mostrando sólo la información verdaderamente útil.

Es entonces que en el año 2001 se acuña el concepto de "Web Semántica", impulsado por el propio inventor de la Web y presidente del consorcio W3C, Tim

Berners-Lee, con el objetivo de que las computadoras puedan comprender los contenidos de la Web, interpretarlos y además utilizarlos. Pero no es hasta 2006 que se establece el término de Web 3.0 en un artículo de Jeffrey Zeldman crítico de la Web 2.0, en donde establece la incorporación de nuevas tecnologías a los motores de búsqueda que permitiera agregar significado semántico al contenido de la Web. La evolución de la web se presenta en figura 1, en donde, además, se establecen los puntos importantes de la web semántica.



Fuente: <https://goo.gl/mRttqX>

Figura 1 Evolución de la Web.

Se menciona en la página "lawebsemantica.weebly.com" que la web "ha caído en un desorden en cuanto a programación y estructura de los documentos, lo que no permite utilizar eficientemente el potencial de los sistemas informáticos.", la web semántica permitirá mejorar las búsquedas en Internet por medio de una estructura más general y sólida de la información en Internet, como menciona la misma página "es un proyecto que pretende configurar información legible no sólo para los usuarios, sino también para las máquinas. Son nuevas formas para estructurar la información, con el propósito de que las computadoras "entiendan" los contenidos que alojan, a fin de estructurar y usar la información con base en una evaluación de tipo semántico. Con esta estructuración y manejo semánticos de la información, los datos contenidos en la Web serán "comprendidos" por las computadoras, así como "almacenados, procesados y reutilizados automáticamente por las mismas" [Pascual, Valdés, y Gómez, s.f.].

Contextualizando específicamente sobre el tema encontraremos que la web semántica es un área de influencia en la Inteligencia Artificial y las tecnologías web, que propone introducir descripciones explícitas sobre el significado de los

recursos, para permitir que las propias máquinas tengan un nivel de comprensión de la web suficiente como para hacerse cargo de una parte del trabajo que actualmente es realizada manualmente por los usuarios que navegan e interactúan con la web.

La web semántica es un término que ha ganado camino por la popularidad de la inteligencia artificial que no solo se enfoca en imitar el sistema cognitivo del ser humano sino además es funcional para muchas otras aplicaciones. Dentro de la web, la necesidad de enfocarse al resultado de información útil ha permitido establecer la idea de implementar la semántica a las búsquedas y dejar de lado las búsquedas basadas puramente en sintaxis. La estructura de las búsquedas basadas en sintaxis tiene la desventaja que al momento de realizar la consulta pueden tomar cualquiera de los elementos que conforman esa estructura e interpretarlo de acuerdo con esa palabra, en figura 2 muestra una consulta realizada en el buscador "Google".



Fuente: <https://www.google.com.mx>

Figura 2 Consulta en la web sintáctica.

Es necesario comprender en qué consiste la web semántica, y ésta, contiene una serie de elementos que la conforman y que son necesarios establecer para su funcionamiento. Dentro de la web semántica existen tres visiones principales:

- Recuperación de la información: Es el texto o los enlaces que se generan al realizar una consulta.
- Base de datos: Son los objetos y entidades que están relacionados con los resultados de la búsqueda.
- Inteligencia Artificial: Son los conceptos o conocimiento que se genera a través de la consulta.

La Web semántica trabaja bajo esquemas conceptuales que sirven para darle un significado semántico a la Web y reglas que sirven para definir objetos y relaciones que existen entre ellos.

## **2. Metodología**

La web semántica implica el análisis del significado de las búsquedas en la web por lo que se seleccionó el método cualitativo que permitirá el estudio de los significados complejos y se enfocará en la problemática que se plantea. A su vez el método de investigación cualitativo contiene una serie de técnicas que ayudan a la recolección de datos, de los cuales se utilizará principalmente la recopilación para obtener la información necesaria para realizar la investigación, las técnicas utilizadas en esta investigación se enumeran a continuación:

- Observación Se realizó una observación porque el estudio se basó en el tema a tratar, en el que se expresó el problema y se llevó a la investigación. ¿Cómo es que los motores de búsqueda nos ayudan a encontrar el tema de interés que se desea buscar?
- Recopilación o investigación documental La investigación documental fue primordial en la recopilación de información respecto a la web semántica, el auge que ha tenido en los últimos años la inteligencia artificial ha permitido explorar diferentes caminos para su aplicación, y por tanto adquirir mayor conocimiento respecto a la web semántica.

En primera instancia, se realizó la investigación de la web semántica en diversos artículos encontrados en Internet, se recopiló información relevante respecto al tema, se analizó y se seleccionó aquella que se consideró de mayor importancia y que fuera totalmente útil para la explicación del tema a tratar.

Una vertiente del método cualitativo es el método inductivo, este método se encarga de generalizar un conocimiento obtenido, en una o varias ocasiones similares, que pueden presentarse en el futuro o en otras latitudes, es decir, van de lo particular a lo general. El método cualitativo subraya la validez de los hechos. En la presente investigación es importante que los resultados se encuentren basados en hechos, para considerar que la web semántica es precisamente eso, una web semántica.

### **3. Resultados**

De acuerdo con el Instituto de Investigación en Inteligencia Artificial (CSIC), se entiende por web semántica a una forma de organizar el contenido en la web que mejore la cooperación entre computadoras y seres humanos [Schorlemmer, 2012]. Se menciona en el artículo "Estado actual de la Web 3.0 o Web Semántica" que "La información que contienen actualmente las páginas de Internet, carecen de una estructura preestablecida, lo que impide que las máquinas sean capaces de procesar la información, como si se tratase de un ser humano, es decir, comprendiendo su significado" [Salazar, 2011].

La web semántica nos ofrece la organización de la gran cantidad de Información suelta, redundante y de calidad dudosa, reducir los costos y tiempos que invertimos en localizar información, establecer reglas para integrar información con diferentes formatos y, resolver problemas de interoperabilidad entre diversos dispositivos y plataformas disponibles para acceder a la Web [Salazar, 2011].

Pero ¿realmente entendemos lo que es la web semántica? El editor y conferencista Dolores Reig de la página web "El caparazón blog" nos explica que "La web semántica es aquella que se construye en un lenguaje que los ordenadores pueden "entender". Todo esto se define para la experiencia de usuario más cómoda y satisfactoria [Dolores, 2009].

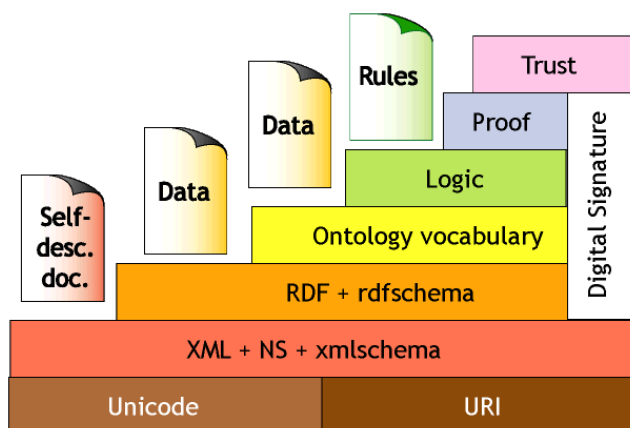
En 2001, se dio a conocer la propuesta de una nueva forma de organizar el contenido en la Red que desencadenaría una avalancha de posibilidades, y, en consecuencia, revolucionaría Internet. Su objetivo es permitir que los datos almacenados en la Web puedan ser procesados por las máquinas de manera inteligente, facilitando a las personas la búsqueda, integración y análisis de la información disponible en la Web. A partir de la definición en 2001, se han desarrollado diferentes tecnologías para la implementación de la misma, en ciudades inteligentes, telemedicina, investigación y colaboración científica, entre otras [Vázquez, Morales, Nájera, y Vieyra, 2014].

En el artículo "La Web Semántica" de Codina & Rovira [2006] menciona que los tres grandes rasgos que debe contener la web para considerarse web semántica son en primera instancia "la codificación de páginas en la cual las etiquetas tengan carga semántica", el segundo rasgo indica que es necesario "aportar descripciones (metadatos) de las páginas y sitios web con un formato que sea compatible con la estructura general de la Web y con diversas categorías de páginas e interoperable entre distintos sistemas informáticos. De este se ocupa la norma *Resource Description Language* (RDF)" RDF según el consorcio W3C es "un modelo" estándar para el intercambio de datos en la Web". El último y tercer rasgo mantiene que es necesario "un sistema de ontologías que permitan especificar conceptos de los diversos dominios del conocimiento mediante el uso de un lenguaje fuertemente basado en lógica simbólica y susceptible, por tanto, de ser eventualmente interpretado por una computadora. De este aspecto se ocupa el denominado *Web Ontology Language* (OWL)." como se menciona en W3C.

Berners-Lee estableció en su artículo "*The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*" que la web semántica estaba representada por una serie de escalones que contienen tanto los lenguajes que muestran el contenido de la web como metadatos y ontologías, esta arquitectura se muestra en la figura 3.

La web semántica nos ofrece algunas ventajas que permiten perfeccionar las búsquedas realizadas en Internet y que permitirán obtener la información útil después de realizar una búsqueda en la web:

- Organización de la información.
- Búsquedas mejoradas basadas en significado.
- Permite la gestión de conocimiento.



Fuente: <https://goo.gl/j5Ed6w>

Figura 3 La Web semántica vista como un sistema de capas Tim Berners-Lee.

Así como contiene ventajas la web semántica existen imprevistos que pueden ocasionar que este tipo de búsquedas sean complicadas de realizar bajo el contexto que se establece, a continuación, se enumeran las desventajas que puede presentar de la web semántica al intentar ser implementada en toda la web:

- Los documentos encontrados en Internet deben ser adaptados para poder ser procesados de manera semántica.
- Establecer los mismos estándares semánticos.

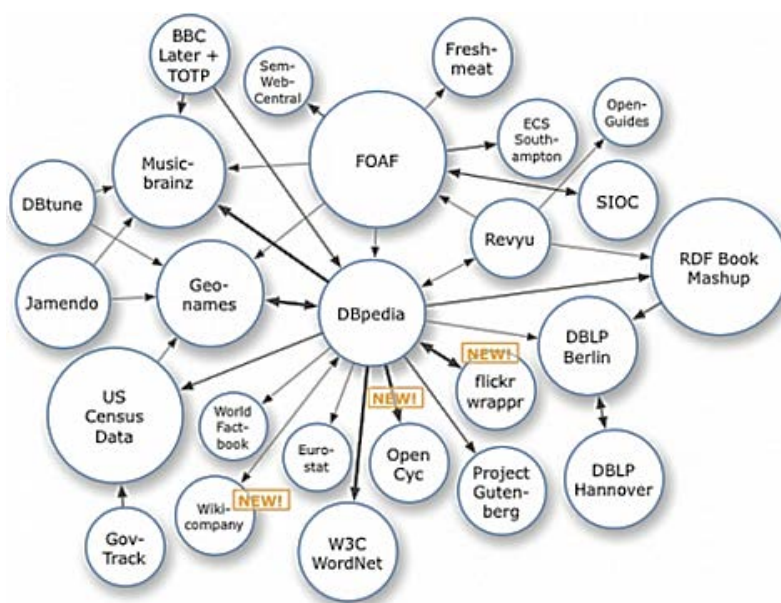
Para tener en toda la web la adaptación necesaria para mejorar las búsquedas es necesario establecer una serie de normas que permitan que la estructura de la información se encuentre siempre de la misma manera en todas las páginas en la web y que todo aquello relevante para el motor de búsqueda se encuentre bien establecido para poder ser identificado con mayor sencillez.

Una de las herramientas que permiten a la Web semántica tener el resultado esperado de las consultas realizadas en el motor de búsqueda son los datos enlazados, la guía breve de Linked Data menciona que "Los Datos Enlazados es la forma que tiene la Web Semántica de vincular los distintos datos que están



distribuidos en la Web, de forma que se referencian de la misma forma que lo hacen los enlaces de las páginas web. Linked Data permite construir la Web de los datos, una gran base de datos interconectados y distribuidos en la Web. Los datos se vinculan y se exploran de una forma similar a la utilizada para vincular los documentos HTML."

Finalmente, el conjunto de datos interconectados entre sí que representan los datos enlazados se conoce como "Nube de Linked Data" como se muestra en figura 4 [Codina & Rovira, 2006].



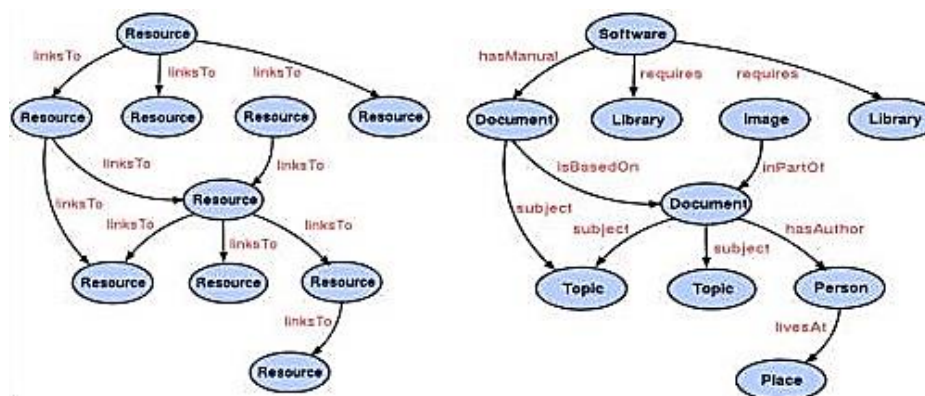
Fuente: <https://goo.gl/tiqRkW>

Figura 4 Nube de Linked Data.

La creación de un ecosistema de aplicaciones y servicios web inteligentes, ubicuos y conscientes de sí mismos y del entorno; no sólo consumidores de la información disponible, sino publicadores de una nueva información proveniente no solo de la web sino también del mundo real. Este ecosistema converge de manera natural con los esfuerzos hechos en diversos frentes (científicos, tecnológicos, políticos, etc.) para avanzar hacia el Internet del Futuro [Vázquez, Morales, Nájera & Vieyra, 2014], figura 5.

El futuro no está muy lejos, recordemos las diferentes propuestas de proyectos de avance tecnológico, así como ver hacia el pasado y explorar las diversas

evoluciones que ha logrado la tecnología a través de la historia. La web semántica solo es un paso más hacia el éxito y evolución de este siglo.



Fuente: <https://goo.gl/wxtL7o>

Figura 5 Comparativa de Web Actual y Web Semántica.

#### 4. Discusión

Alguna vez te has preguntado el por qué anteriormente cuando se buscaba algún tema de interés en un motor de búsqueda daba diferentes entradas a lo que realmente se buscaba. Por ejemplo, si se escribía "Artículo sobre Gabriel García Márquez", el resultado daba entradas como "Artículos de Gabriel García Márquez". Esto es muy sencillo de explicar, dado que los buscadores se basan principalmente en búsquedas sintácticas, es decir, no encontraban el sentido de las palabras para realizar la búsqueda, esto limitaba la capacidad de búsqueda. Actualmente la mayoría de los buscadores utilizan la web sintáctica, es decir, el motor de búsqueda se basa en términos, la información que se arroja es amplia, prácticamente toda la información disponible de todo el mundo que contenga esa palabra o serie de palabras son mostradas, es por ello por lo que al realizar una consulta en un buscador este es capaz de mostrar una cantidad impresionante de información. En cambio, la web semántica se basa principalmente en el significado de las palabras, los servidores son capaces de realizar consultas inteligentes, lo que permite obtener un resultado más congruente de lo que desea. Una de las más grandes ventajas es que no necesita la intervención humana para interpretar la información.

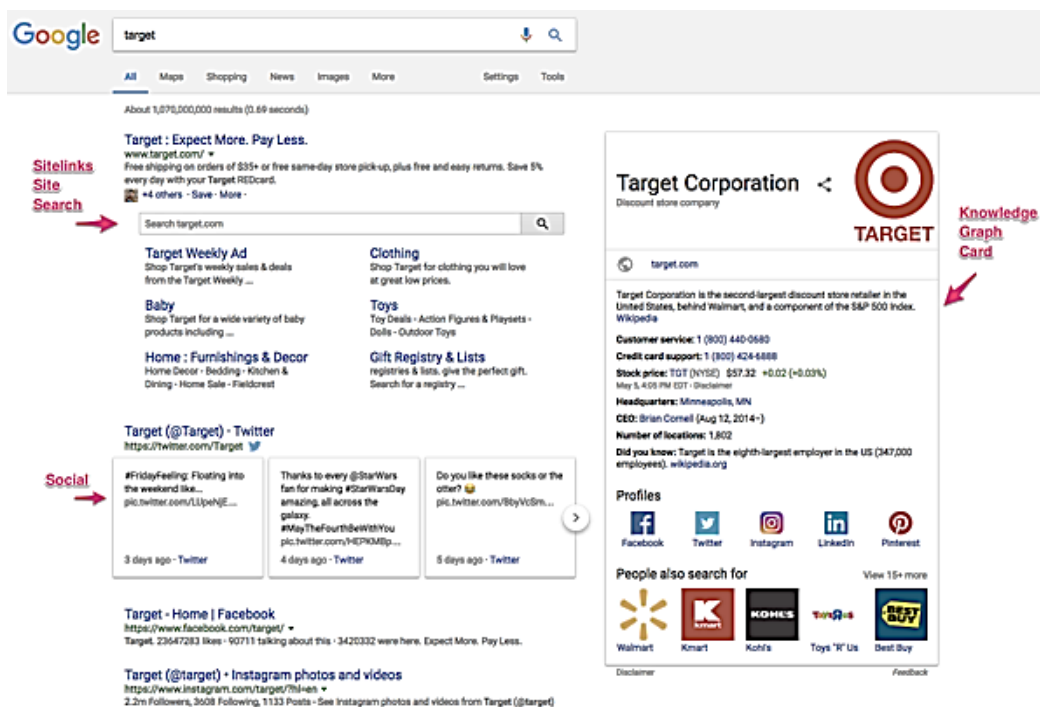
En general, hoy en día casi todo se relaciona con la web, ya sea de una forma u otra, hace viable la búsqueda y más que nada el encontrar resultados pertinentes, casi como una enciclopedia que contiene N cantidad de conocimientos. Lo cual es favorable ya que permite hacer trabajos, compras, ventas, búsquedas, consultas, etc., desde un sitio seguro.

La web 3.0 es el nombre que se le da en general al cambio evolutivo de la web este incluye la web semántica. Si bien al hablar de la web 3.0 no se refiere a que es lo mismo que la web semántica porque ésta va incluida en el término. Según Claudio Gutiérrez y Carlos Hurtado mencionan en su artículo "La Web Semántica: Realidades y Perspectivas" que las asociaciones semánticas son "Relaciones complejas que capturan conectividad y similaridad de entidades en una base de conocimientos" [Gutiérrez & Hurtado, s.f.].

En mayo de 2012 Google lanzó su nuevo algoritmo de búsqueda semántica implementada en una API fue llamado Knowledge Graph y se basa principalmente en la búsqueda dentro de una red semántica. Este algoritmo permite realizar búsquedas ambiguas ya que Google puede entender a lo que se refiere y sugiere una lista de opciones que pueden ser de ayuda. Dentro de Knowledge Graph existe una cantidad considerable de entradas relacionadas con el mundo real, según la página <https://developers.google.com/knowledge-graph/> cada entrada forma un nodo para Knowledge Graph. A continuación, se muestra una serie de entidades que se encuentran dentro de esta API: Book, BookSeries, EducationalOrganization, Event, GovernmentOrganization, LocalBusiness, Movie, MovieSeries, MusicAlbum, MusicGroup, MusicRecording, Organization, Periodical, Person, Place, SportsTeam, TVEpisode, TVSeries, VideoGame, VideoGameSeries, WebSite.

"Esta API contiene datos sobre personas, cosas y lugares y cómo estas entidades están relacionadas entre sí. Esta red ayuda a mejorar la relevancia de búsqueda." [Toonen, 2017]. Según la página "What is Google's Knowledge Graph?" Un ejemplo de Knowledge Graph es, si se realiza una búsqueda de una película, mostrará los carteles, reseñas y horarios de las funciones de los cines más cercanos a tu ubicación. Su objetivo principal es brindar una respuesta correcta a

todas las preguntas que el usuario realice, esté presenta los resultados de manera más completa ya que existen conexiones más amplias entre los datos. Esta API recopila gran información y analiza cantidades masivas de datos sobre personas, lugares, cosas y hechos, así mismo desarrolla formas de presentar ese conocimiento de una forma más accesible. En la figura 6 se muestra una consulta realizada en Google con este algoritmo de búsqueda semántica.



Fuente: <https://goo.gl/nyRygg>

Figura 6 Información que se muestra al realizar una búsqueda semántica.

## 5. Bibliografía y Referencias

- [1] Castells, P. (2006). La web semántica., de Escuela Politécnica Superior Universidad Autónoma de Madrid: <https://goo.gl/CkgeyR>.
- [2] Codina, L. & Rovira, C. (2006). *La Web Semántica*. Universidad Pompeu Fraba, Departamento de Periodismo y de Comunicación Audiovisual, Instituto Universitario de Lingüística Aplicada: <https://goo.gl/wLLq9i>.
- [3] Dolores, R. (2009). Entrevista web semántica 3.0. para la revista Computing. Del Blog y Academia El Caparazón: <https://goo.gl/HDbasV>.

- [4] Gutiérrez, C. & Hurtado, C. (s.f.). Web semántica: Realidades y perspectivas. Del Centro de Investigación de la Web, Departamento de ciencias y computación, Universidad de Chile: <https://goo.gl/4Lb2HZ>.
- [5] Intefblog. (2014). Web Semántica: un mundo de posibilidades. Del Instituto Nacional de Tecnologías Educativas y de Formación de Profesorado: <https://goo.gl/9BPDts>.
- [6] Lamarca, M.J. (2013). Hipertexto: El nuevo concepto de documento en la cultura de la imagen. De la Universidad Complutense de Madrid: <https://goo.gl/7RLavV>.
- [7] López, N. & Sandoval, I. (s.f.). Métodos y técnicas de investigación cuantitativa y cualitativa. Del Sistema de Universidad virtual, Universidad de Guadalajara: <https://goo.gl/3Q6XR5>.
- [8] Pascual, I. Valdés, O. y Gómez, E. (s.f.). La Web Semántica: <https://goo.gl/yKXZiw>.
- [9] Salazar, J. (2011). Estado actual de la Web 3.0 o Web Semántica. Revista UNAM: <https://goo.gl/UVMRA9>.
- [10] Schorlemmer, M. (2012). Diez años construyendo una web semántica., de Instituto de Investigación en Inteligencia Artificial: <https://goo.gl/p32UQL>.
- [11] Toonen, E. (2017). What is Google's Knowledge Graph? <https://goo.gl/pHHp6m>
- [12] Vázquez, B. Morales, E. Nájera, K, & Vieyra S. (2014). Web Semántica: La evolución de una web con significado. De SG Buzz: <https://goo.gl/LVzpnA>.
- [13] W3C. (s.f.). Guía breve de web semántica., de W3C: <https://goo.gl/E5vsve>.